複数時系列中の類似セグメント高速探索法

杉山雅英[†1]

特徴ベクトルの複数の時系列中に共通に含まれる類似セグメント探索問題を定式化し，高次元時刻空間における枝刈/検出判定のための距離計算をスキップできる領域が超菱形（ℓ_1 球）になることを示し，その超菱形の半径の評価式を導出する．さらに類似セグメント探索問題の解法として再帰的菱形分割探索法（RDDS: Recursive Diamond Division Search 法）を提案し，2 つおよび 3 つの時系列中の類似セグメント探索問題に適用しその有効性を評価する．実験資料として 1 時間長の音楽リクエスト番組を用いる．2 つの時系列に含まれる類似セグメント探索実験の結果から AS 反復探索法に比べて約 163 倍の高速化が得られること，30 分長の 3 つの時系列に含まれる類似セグメント探索実験の結果から全数探索法に対して 4.910×10^6 倍高速化されることを示す．

An Efficient Similar Segment Search Algorithm in Multiple Time Series

Masahide Sugiyama[†1]

This paper formulates a similar segment search problem in multiple time series of the feature vectors and proves that skip region for rejection/detection is equal to a hyper diamond shape (ℓ_1 ball) in the time index space and the upper bound of the ball radius is derived. In order to solve the search problem, this paper proposes a new algorithm: RDDS (Recursive Diamond Division Search), and implemented and evaluated RDDS in two and three time series problem. The experiment results show that RDDS runs about 163 times faster than repeated AS algorithm in two time series and 4.910×10^6 times faster than the exhaustive search algorithm in three time series.

1. まえがき

ハードディスクの低価格化と大容量化，デジタルカメラ・カメラ付きの携帯電話やデジタルビデオカメラの普及，さらに高速インターネットの利用により，音声・画像・ビデオなどのマルチメディア情報が容易に蓄積・流通できるようになってきた．蓄積されたマルチメディアデータに検索のための情報が付加されているとは限らないのでデータ検索技術はますます必要とされている．また作者からの許諾なしのマルチメディアデータの再利用は深刻な問題であり，与えられたマルチメディアデータ間の類似部分や共通部分を検出する技術は著作権の保護のために有用である．一方，異なる地点や時刻で得られた観測データに含まれる共通・類似部分を検出することで一見無関係な現象に含まれる共通の現象を新たに発見することや，1 つのデータに含まれる類似部分を見つけ出すことでデータの持つ構造を解析できる．複数のマルチメディアデータや観測データは複数の時系列（特徴ベクトルの系列）と見なせるので複数の時系列に含まれる類似部分・共通部分を高速に検出・探索する手法が必要となる．

時系列に含まれるクエリ高速探索手法として，柏野らは Active 探索法（以下では AS 法と略する）を提案しその有効性を示した[1)]．AS 法の基本原理はある時刻で計算した類似度（距離）とその近傍での類似度の値との差の上限を用いて類似度計算をスキップ（削減）することができる性質を利用することである．特徴ベクトルを有限個の代表ベクトルで表現することで特徴ベクトルの一定長時系列（セグメントと呼ぶ）を代表ベクトルのヒストグラムで表現する．類似度である正規化ヒストグラム間の重なり度に対して時刻軸での近傍での類似度の差の上限を導き出した．一方，我々は AS 法の導出で用いられる類似度を距離に置き換えることで近傍での類似度の差の上限の導出を距離の持つ三角不等式に帰着できることを示し，AS 法の原理の明確化を行った[3)]．

一方，西村ら[4),5)] は 2 つの時系列に含まれる類似セグメント探索問題に対して AS 法のスキップ幅が 2 次元空間において菱形になることに着目しスキップ領

†1 会津大学大学院コンピュータ理工学研究科
The University of Aizu

域を 2 次元空間において接続することで AS 法を 2 次元に拡張し，RIFAS 法（Reference Interval Free Active Search）を導出し AS 法を反復して適用する手法に比べて 20–40 倍程度の高速化を実現した．AS 法をクエリ軸方向に拡張する手法であるので処理は逐次的であり，菱形の中心から見て上半面，すなわち菱形の 2 分の 1 しか利用できていない．また菱形の交点計算および交点の s 軸での最小値を求めるので処理が複雑である．さらに，たとえば 3 次元空間における正 8 面体の接続は容易ではないので RIFAS 法は 3 つ以上の時系列に含まれる類似セグメント探索への拡張は難しい．柏野ら[6] も西村らと同一の手法で 2 つの時系列中の部分信号区間の高速抽出法を提案し AS 法を反復する方法に比べて 10 倍程度高速化できることを報告している．

本論文では任意個数の時系列に共通に含まれる類似セグメント探索問題を定式化し，高次元時刻空間におけるスキップ半径の評価式を導出し探索問題の解法として RDDS 法を提案する．探索評価実験により AS 法および RIFAS 法と比較し RDDS 法の有効性を明らかにする[7]．

本論文は以下のように構成されている．2 章で AS 法および RIFAS 法について述べる．3 章で複数時系列中の類似セグメント探索問題の定式化とその解法の提案を行い，4 章で提案した手法の評価を行う．

2. AS 法と RIFAS 法

2.1 出現確率の時系列の性質

音声やビデオなどの特徴ベクトルの時系列 $\boldsymbol{v}_t$（$t = 0, \ldots, \hat{T}-1$）を M 次元空間の部分空間 $\boldsymbol{U}_1 = \{\boldsymbol{x} = (x_m) \in R^M \mid x_m \geq 0, ||\boldsymbol{x}||_1 = 1\}$ のベクトル $\boldsymbol{r}_t$（たとえばベクトル量子化による出現符号ベクトル）に変換する．ここで $||*||_1$ は ℓ_1 ノルムであり，L 時刻分の $\boldsymbol{r}_t$ の相加平均を $\boldsymbol{p}_t$ とする．L はセグメント長に対応する．$\boldsymbol{U}_1$ は凸である[8] ので $\boldsymbol{p}_t \in \boldsymbol{U}_1$（$t \in [0, T)$, $T = \hat{T} - L$）となる．この時系列を $\boldsymbol{P} = (\boldsymbol{p}_t)$ と表す．任意の時系列 $\boldsymbol{P}$ の任意の時刻のベクトル $\boldsymbol{p}_t \in \boldsymbol{P}$ および任意のベクトル $\boldsymbol{q} \in \boldsymbol{U}_1$，整数 n および ℓ_p 距離（$p \geq 1$）に対して式 (1) が成り立つ[9]．

$$| d_p(\boldsymbol{p}_{t+n}, \boldsymbol{q}) - d_p(\boldsymbol{p}_t, \boldsymbol{q}) | \leq \frac{2^{\frac{1}{p}}|n|}{L}. \qquad (1)$$

2.2 AS 法（Active 探索法）とそのスキップ幅

時系列 $\boldsymbol{P}$ に含まれるクエリ探索問題は以下で定式化される．

時系列に含まれるクエリ探索問題

時系列 $\boldsymbol{P} \subset \boldsymbol{U}_1$ に対して $\boldsymbol{q} \in \boldsymbol{U}_1$ をクエリ，$\theta \geq 0$ を探索閾値とし，不等式 (2) を満たす時刻 t を見つけること．

$$d_p(\boldsymbol{p}_t, \boldsymbol{q}) \leq \theta. \qquad (2)$$

幾何学的に述べれば $\boldsymbol{q}$ を中心とする半径 θ の球に属する $\boldsymbol{p}_t$ の時刻 t を求める問題となる．全数探索法の距離計算回数は時系列の長さ T に等しい．この高速化問題は距離計算回数の削減で実現される[⋆1]．不等式 (1) から時刻 t での距離 $d_p(\boldsymbol{p}_t, \boldsymbol{q})$ と探索閾値 θ で式 (3) で定まる値

$$R = \frac{L}{2^{\frac{1}{p}}} |\theta - d_p(\boldsymbol{p}_t, \boldsymbol{q})|, \qquad (3)$$

に対して $\forall |n| < R$ を満たすとき，式 (4) が成り立つ．

$$\begin{cases} d_p(\boldsymbol{p}_t, \boldsymbol{q}) > \theta & \Longrightarrow d_p(\boldsymbol{p}_{t+n}, \boldsymbol{q}) > \theta, \\ d_p(\boldsymbol{p}_t, \boldsymbol{q}) \leq \theta & \Longrightarrow d_p(\boldsymbol{p}_{t+n}, \boldsymbol{q}) \leq \theta. \end{cases} \qquad (4)$$

ここで R をスキップ幅と呼ぶ．AS 法の動作原理を**図 1** に示す．時刻 t でクエリとの距離 $d_p(\boldsymbol{p}_t, \boldsymbol{q})$ を計算し，$\boldsymbol{p}_t$ が探索球の内側（外側）であれば $n < R$ を満たす時刻までは式 (4) より $\boldsymbol{p}_{t+n}$ が探索球の内側（外側）にあることになるのでその距離 $d_p(\boldsymbol{p}_{t+n}, \boldsymbol{q})$ を計算する必要がない．次に距離を計算しなければならない時刻は $t+R$ となり，$t+1, t+2, \ldots, t+R-1$ での距離計算をスキップできることになる．

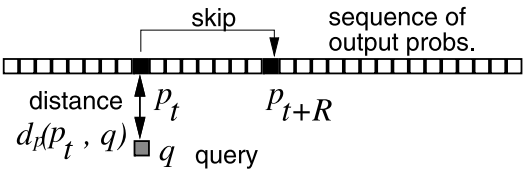


図 1 AS 法のスキップ幅
Fig. 1 Skip width in AS algorithm.

文献 1) の AS 法では式 (2) を満たす場合のスキップ幅を 1 とし全数探索に切り替え，式 (2) を満たすと同時に $d_p(\boldsymbol{p}_t, \boldsymbol{q})$ が最小，すなわち最近傍となる時刻を探索する問題を扱っている．本論文の AS 法の定式化では式 (2) を満たさない，すなわち探索球の外側の場合（outside AS 法）だけでなく探索球の内側でもスキップ（inside AS 法）を用い探索処理を高速化している[2]．したがって，2.3 節で述べる AS 反復法においても同様に探索球の外側および内側においてスキップ処理を行うことで処理を高速化している．

2.3 RIFAS 法

RIFAS 法は 2 つの時系列に含まれる類似部分を高

⋆1 距離に基づいて導出法を述べるが類似度に基づいた導出も同様である．

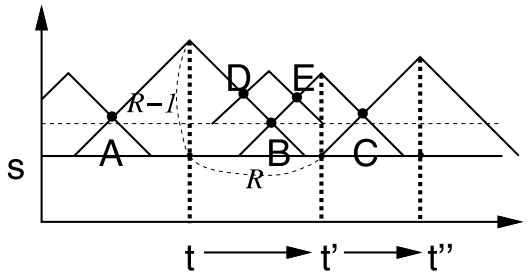


図 2 RIFAS 法とスキップ領域
Fig. 2 RIFAS algorithm and its skip region.

速探索する手法である．AS 法のスキップ幅が 2 次元空間において菱形になることに着目しスキップ領域を 2 次元空間において接続することで AS 法を 2 次元に拡張した方法である．

2 つの時系列に含まれる類似部分探索問題

2 つの時系列 $\boldsymbol{P}=(\boldsymbol{p}_t)$，$\boldsymbol{Q}=(\boldsymbol{q}_s)\subset \boldsymbol{U}_1$ に対して，$\theta\geq 0$ を探索閾値とし，不等式 (5) を満たす時刻の組合せ (t,s) を見つけること．

$$d(\boldsymbol{p}_t,\boldsymbol{q}_s)\leq\theta. \tag{5}$$

時刻 s を $s=0$ とし，$\boldsymbol{q}_s$ を AS 法におけるクエリとして $d(\boldsymbol{p}_t,\boldsymbol{q}_s)\leq\theta$ を求める t を満たす問題を考え，t 軸に AS 法を適用して式 (3) のスキップ幅 R を求めることで高速化可能である．これを AS 反復法と呼ぶことにする．このとき，**図 2** に示すように $\boldsymbol{p}_t$ をクエリと見なすと s 軸に関してもスキップ幅 R だけスキップ可能である．したがって (t,s) 平面でスキップ可能な領域は (t,s) を中心とする菱形となる．t 軸に沿って t' にスキップするごとに菱形のスキップ領域が定まるので，重複する菱形の交点 (t,s) を算出し交点集合 { A, B, C } の中の s 軸における最小値を与える点 B で距離を計算し，新たに計算された菱形との交点 D，E を求める．この処理を繰り返すことで距離計算を削減できる．

従来の RIFAS 法では AS 法と同様に閾値条件を満たす入力データ軸での最近傍時刻の探索問題を扱っているが，本論文では時系列に含まれる閾値条件を満たす時刻の組と類似部分を定義しその探索問題を扱っている．

3. 複数時系列中の類似セグメント探索問題

3.1 探索問題の定式化とスキップ半径

複数（I 個）の時系列 $\boldsymbol{P}^{(i)}\subset\boldsymbol{U}_1$（$i\in[1,I]$）の時刻を t_i で，i，j 番目の時系列の時刻 t_i，t_j のベクトル $\boldsymbol{p}_{t_i}^{(i)}$，$\boldsymbol{p}_{t_j}^{(j)}$ 間の距離 $d_p(\boldsymbol{p}_{t_i}^{(i)},\boldsymbol{p}_{t_j}^{(j)})$ を $d(t_i,t_j)$ と表す．時刻ベクトル $\boldsymbol{t}=(t_1,t_2,\ldots,t_I)$ における時系列の距離を式 (6) で定義する．

$$d(\boldsymbol{t})=\sum_{i<j}d(t_i,t_j). \tag{6}$$

距離 $d(t_i,t_j)$ の対称性から和は $i<j$ の時刻の組とし，組数は ${}_IC_2=I(I-1)/2$ で $d(t_i,t_j)\leq 2^{\frac{1}{p}}$ なので $d(\boldsymbol{t})$ の値域は式 (7) で与えられる．

$$0\leq d(\boldsymbol{t})\leq I(I-1)2^{\frac{1}{p}-1}. \tag{7}$$

類似セグメント探索問題を以下で定式化する．

複数時系列類似セグメント探索問題

I 個の時系列 $\boldsymbol{P}^{(i)}$ に対して $\theta\geq 0$ を探索閾値とし，不等式 (8) を満たす時刻ベクトル $\boldsymbol{t}$ を見つけること．

$$d(\boldsymbol{t})\leq\theta. \tag{8}$$

式 (8) を満たす時刻ベクトル $\boldsymbol{t}=(t_1,t_2,\ldots,t_I)$ が I 個の時系列に含まれる類似セグメントの時刻を示す．$d(\boldsymbol{t})=0$ であれば $\boldsymbol{p}_{t_i}^{(i)}=\boldsymbol{p}_{t_j}^{(j)}(\forall i,j)$ となる．すべての時刻の組 $\boldsymbol{t}$ に対して距離 $d(\boldsymbol{t})$ を求めれば式 (8) を満たす時刻を検出することができるがその距離計算回数は時系列の長さ T_i の積 $T=\prod_i T_i$ となり膨大な数となる．したがって探索の高速化問題は式 (6) の距離計算回数を削減することで実現される．式 (7) の $d(\boldsymbol{t})$ の値域から θ の値域は式 (9) で与えられる．

$$0\leq\theta\leq I(I-1)2^{\frac{1}{p}-1}. \tag{9}$$

式 (1) に対応する不等式 (10)，(11) が成り立つ．

性質 1 時刻 $\boldsymbol{t}=(t_i)$，$\boldsymbol{t}+\boldsymbol{n}=(t_i+n_i)$ における距離 $d(\boldsymbol{t})$ は以下の不等式を満たす．

$$|d(\boldsymbol{t}+\boldsymbol{n})-d(\boldsymbol{t})|\leq 2^{\frac{1}{p}}(I-1)\sum_i\frac{|n_i|}{L_i}. \tag{10}$$

ここで L_i は平均化窓長（セグメント長）であり，$L_i=L$ の場合には以下で与えられる．

$$|d(\boldsymbol{t}+\boldsymbol{n})-d(\boldsymbol{t})|\leq\frac{2^{\frac{1}{p}}(I-1)}{L}||\boldsymbol{n}||_1. \tag{11}$$

ここで $||\boldsymbol{n}||_1=\sum_i|n_i|$ $(\boldsymbol{n}=(n_i))$ である．
証明を付録 A.1 に示す．式 (10)，(11) は式 (1) の多次元時刻への拡張である．以下では $L_i=L$ に限定して述べる．

性質 2 距離 $d(\boldsymbol{t})$ と探索閾値 θ で定まる値

$$R=\frac{L}{2^{\frac{1}{p}}(I-1)}|\theta-d(\boldsymbol{t})|, \tag{12}$$

に対して，時刻ベクトル $\boldsymbol{n}$ が不等式 (13) を満たすとする．

$$||\boldsymbol{n}||_1<R. \tag{13}$$

このとき，$d(\boldsymbol{t})>\theta$ であれば $d(\boldsymbol{t}+\boldsymbol{n})>\theta$ であり，$d(\boldsymbol{t})<\theta$ であれば $d(\boldsymbol{t}+\boldsymbol{n})<\theta$ となる．

証明を付録 A.2 に示す．式 (13) で等号を含めると

$d(\boldsymbol{t}+\boldsymbol{n}) \leq \theta$ 等となる．多次元時刻 $\boldsymbol{t}$ とすることで式 (3) のスキップ幅 R は I 次元時刻空間における ℓ_1 ノルムに対する不等式を与える．式 (3) の R は式 (12) の $I=2$ の場合に対応する．式 (12) で定義される R をスキップ半径と呼ぶことにする．式 (13) を満たす時刻ベクトル $\boldsymbol{n}$ は ℓ_1 ノルムの球で I 次元空間の超菱形領域（1 次元では区間，2 次元では菱形，3 次元では正 8 面体）となる．式 (7)，式 (9) の $d(\boldsymbol{t})$，θ の値域からスキップ半径は不等式 (14) を満たす．

$$R \leq \frac{IL}{2}. \tag{14}$$

3.2 探索とスキップ半径

探索時は時刻 $\boldsymbol{t}$ での距離 $d(\boldsymbol{t})$ を計算し式 (8) を判定する．$d(\boldsymbol{t}) \leq \theta$ であれば W（$< R$）を半径とする ℓ_1 球 $B_1(\boldsymbol{t}, W) = \{\boldsymbol{x} \in R^I \mid ||\boldsymbol{x}-\boldsymbol{t}||_1 \leq W\}$ 内のすべての時刻 $\boldsymbol{t}+\boldsymbol{n}$ が式 (8) を満たす．一方，$d(\boldsymbol{t}) > \theta$ であれば球 $B_1(\boldsymbol{t}, W)$ の内部のすべての時刻は式 (8) を満たさないとして枝刈りできる．W が大きければ 1 回の距離計算で多くの時刻の組に対して検出/枝刈りが判定可能であり，距離計算を削減できる．$d(\boldsymbol{t})-\theta$ が 0 に近づく，すなわち検出と枝刈領域の境界付近に $\boldsymbol{t}$ が近づくと R は小さくなり判定の効率が低下する．

3.3 RDDS 法（再帰的菱形分割探索法）

類似セグメント探索問題の解法として RDDS（Recursive Diamond Division Search）法を提案する．時刻領域 $\boldsymbol{T} = [0, T_1) \times [0, T_2) \times \ldots [0, T_I)$ を同一半径 W の ℓ_1 球[⋆1] で被覆する．

$$\boldsymbol{T} \subset \cup_i B_1(\boldsymbol{t}_i, W). \tag{15}$$

球の中心 $\boldsymbol{t}_i$ での距離 $d(\boldsymbol{t}_i)$ からスキップ半径 R を求め，W と R を比較する．いい換えれば式 (12)，(13) から導かれる式 (16) を用いて球 $B_1(\boldsymbol{t}_i, W)$ は枝刈り領域，検出領域，判定不可能のいずれかになる．判定不可能の場合には球を小さな球に分割し新たな球の中心において距離を求め再度判定を行う．

$$\begin{cases} d(\boldsymbol{t}) & > \theta + \dfrac{2^{\frac{1}{p}}(I-1)W}{L}, & \text{枝刈り.} \\ d(\boldsymbol{t}) & < \theta - \dfrac{2^{\frac{1}{p}}(I-1)W}{L}, & \text{検出.} \\ d(\boldsymbol{t}) & \text{上を満たさない}, & \text{球を分割.} \end{cases} \tag{16}$$

2 次元で式 (15) を満たす ℓ_1 球は**図 3** に示すように菱形を用いて構成できるが，3 次元以上の高次元空間では ℓ_∞ 球（超立方体）との以下の関係 (17) を利用して構成する．

⋆1 I 次元空間の ℓ_1 球は正 2^I 面体で頂点数 $2I$，正 I 面体 2^I 個で構成され，ℓ_∞ 球は正 $2I$ 面体で頂点数 2^I，正 $(2I-2)$ 面体 $2I$ 個で構成される．各々互いに双対である．

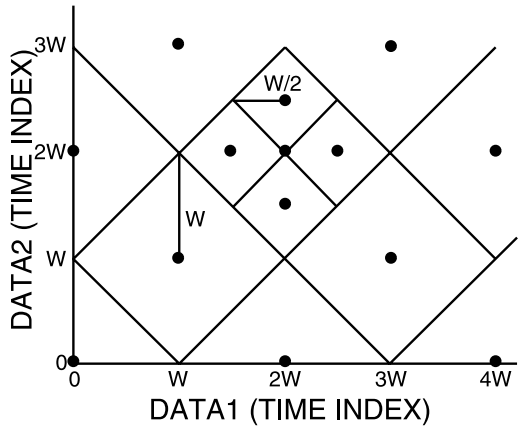


図 3 2 次元空間 RDDS 法における菱形被覆と再帰分割
Fig. 3 Diamond covering and its recursive division in two dimensional RDDS algorithm.

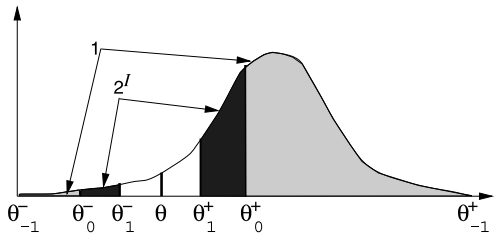


図 4 距離の分布と計算量について
Fig. 4 Distribution of distance values and computation amounts of RDDS.

$$B_\infty(\boldsymbol{t}_i, r) \subset B_1(\boldsymbol{t}_i, Ir). \tag{17}$$

時刻領域 $\boldsymbol{T}$ を被覆する半径 $r = W/I$ の ℓ_∞ 球の中心 $\boldsymbol{t}_i$ を定める．このとき $\boldsymbol{T} \subset \cup_i B_\infty(\boldsymbol{t}_i, W/I) \subset \cup_i B_1(\boldsymbol{t}_i, W)$ を満たす．不等式 (14) から半径は $W \leq (IL)/2$ であるので ℓ_∞ 球の半径の上限は次元によらず式 (18) で定まる．

$$r = W/I \leq L/2. \tag{18}$$

したがって超立方体（ℓ_∞ 球）の 1 辺の上限は L となる．

3.4 RDDS 法の計算量

RDDS 法の計算量（距離計算回数）は時刻空間の被覆に要する球の個数と球の中での平均距離計算回数との積で推定できる．ℓ_1 球の半径を W とし，時系列の長さを T_i とすると式 (17) で作成する ℓ_∞ 球の半径は W/I となり，球の個数は $(\prod_i T_i)/(2W/I)^I$ と推定される．球の中での平均距離計算回数 $a(W)$ は $\boldsymbol{t}$ での距離の値 $\tau = d(\boldsymbol{t})$ の分布関数 $f(\tau)$ によって推定される．球の中での距離計算は**図 4** に示すように距離値 $d(\boldsymbol{t})$ が式 (16) の第 1 もしくは第 2 の条件を満たす場合，すなわち以下で定まる θ_k^+，θ_k^- に対して

$$\begin{cases} \theta_k^+ & = \min\left(\theta + \dfrac{2^{\frac{1}{p}}(I-1)W_k}{L}, 2^{\frac{1}{p}}\right), \\ \theta_k^- & = \max\left(\theta - \dfrac{2^{\frac{1}{p}}(I-1)W_k}{L}, 0\right), \\ \theta_{-1}^- & = 0, \quad \theta_{-1}^+ = 2^{\frac{1}{p}}, \\ W_k & = W/2^k, \quad (k = 0, 1, \ldots, k_{\max}). \end{cases}$$

$d(\boldsymbol{t}) \in [\theta_k^+, \theta_{k-1}^+]$ もしくは $d(\boldsymbol{t}) \in [\theta_{k-1}^-, \theta_k^-]$ を満たす場合には処理は終了するが，第 3 の条件，すなわち，両方とも満たさず $d(\boldsymbol{t}) \in [\theta_k^-, \theta_k^+]$ の場合には球を 2^I 個に分割し再度距離計算を行うことになる．したがって $d(\boldsymbol{t}) \in [\theta_{k-1}^-, \theta_{k-1}^+]$ のとき，$a(W)$ は式 (19) で推定できる．

$$2^{kI}\left[\left(\int_{\theta_{k-1}^-}^{\theta_k^-} f(\tau)d\tau + \int_{\theta_k^+}^{\theta_{k-1}^+} f(\tau)d\tau\right) + 2^I \int_{\theta_k^-}^{\theta_k^+} f(\tau)d\tau\right]. \quad (19)$$

全数探索法と同一の探索結果を与えるためには $W_k = 0$ となるまで再帰分割を行う必要があるが，再帰分割を途中で停止してもほぼ同様の結果が得られ，かつ探索処理を効率化できる．そこで球の半径の最小値 $W_{\min}$ に対して $W_k \geq W_{\min}$ の範囲で探索するとき，$k_{\max} = \log_2(W_0/W_{\min})$（$W_{\min} = 0$ のときは $\log_2 W_{\min} = -1$ とする）であり，この場合 $\theta_k^+ = \theta_k^- = \theta$ とする．$a(W)$ は式 (20) で与えられる．

$$a(W) = \sum_{k=0}^{k_{\max}} 2^{kI}\left(\int_{\theta_{k-1}^-}^{\theta_k^-} f(\tau)d\tau + \int_{\theta_k^+}^{\theta_{k-1}^+} f(\tau)d\tau\right). \quad (20)$$

定義から $1 \leq a(W) \leq 2^{(\log_2 W+1)I}$ である．球の個数を乗じて総計算回数は式 (21) で推定される．

$$C(W) = \frac{\prod_i T_i}{(2W/I)^I}\, a(W). \quad (21)$$

W を小さくすれば第 1 項は大きくなり第 2 項は再帰呼び出しが減るので減少し，ある程度小さくすると $a(W) = 1$ となるので $C(W)$ は急激に増加する．逆に W を大きくすると第 1 項は小さくなるが再帰呼び出し回数が増えるので第 2 項は増加する．したがって $C(W)$ を最小化する $W = W_{\text{opt}}$ が存在し，距離の分布 $f(\tau)$ と閾値 θ で与えられる．

$$W_{\text{opt}} = \arg\min_W C(W). \quad (22)$$

全数探索法の総距離計算回数は $\prod_i T_i$ であるので RDDS 法との総距離計算回数の比 $\rho(W)$ は式 (23) で与えられる．$\rho(W)$ が大きいほど全数探索法に比べて RDDS 法の探索の効率が高いことになる．

$$\rho(W) = \frac{1}{C(W)}\prod_i T_i = \frac{1}{a(W)}\left(\frac{2W}{I}\right)^I. \quad (23)$$

$a(W) \geq 1$ であるので $\rho(W) \leq (\frac{2W}{I})^I$ であり，式 (18) から ρ の上限を与える不等式 (24) を得る．

$$\rho(W) \leq L^I. \quad (24)$$

ρ の上限はセグメント長 L と時系列の個数 I で与えられ，I が大きくなるほど RDDS 法の効率が増大することを示唆している．

3.5 RDDS 法の特長

RDDS 法の特長は以下のとおりである．

(1) 再帰（分割統治法：Divide-and-Conquer）を用いて実現でき実装が容易．
(2) 一般的な原理から導かれ複数時系列探索，多次元クエリ探索問題などの様々な探索問題に適用可能．
(3) 再帰呼び出しを途中でやめることで探索（空間）分解能を容易に変更可能[⋆1]．
(4) 被覆球内で処理が閉じているので球ごとに探索を並列化可能．
(5) 複数の異なる探索閾値（θ）で繰り返し探索を行う場合，球の中心での距離値保持により距離計算処理を効率化可能．
(6) 距離値の分布から探索計算量を推定可能であり，球の初期半径の最適値 W_{opt} の推定や半径の最小値 $W_{\min}$ の効果の評価が可能．

4. RDDS 法の評価

4.1 評価実験条件

評価実験に CampusWave データベース[10] の第 1 回，第 2 回および第 3 回目の収録データを用いた．これは会津若松市内の FM 局の音楽リクエスト番組であり，2 名の女性パーソナリティの対話音声，リクエスト曲，CM 音声などを含み，音声長は各々約 1 時間である．音響分析などの実験条件を**表 1** に示す．$L = 625$（10 秒に対応[⋆2]），VQ 符号帳サイズを $M = 32$ とし第 1 回データから LBG 法で作成した．ℓ_p 距離は $p = 1$ とした[9]．

⋆1 判定に処理時間の要する境界部分で処理を省略することであり，一律に時間間引きをして空間分解能を下げることとは異なる．
⋆2 類似セグメントクエリの検出対象のコマーシャル長に対応している．

表 1 探索実験条件
Table 1 Setup of search experiments.

標本化周波数	16 kHz
窓長	256 点（16 ms）
フレーム更新周期	256 点（16 ms）
窓関数	ハミング窓
高域強調	$(1-0.97\ z^{-1})$
LPC 分析	14
ケプストラム分析	16
音声長	約 1 時間（約 $T = 225{,}000$）
セグメント長	10 秒（$L = 625$ フレーム）
VQ 符号帳サイズ	$M = 32$

4.2 2 つの時系列中の類似セグメント探索

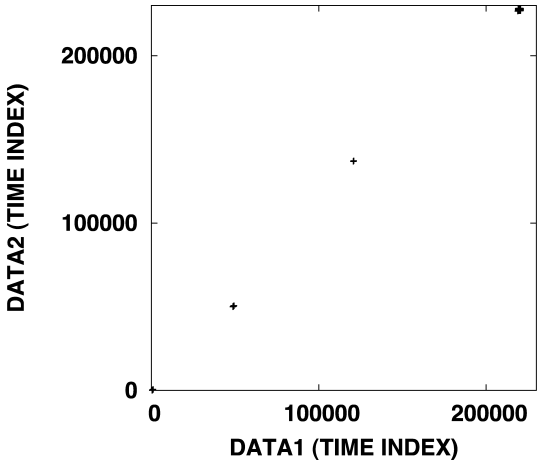
2 つの時系列中の類似セグメント探索問題，すなわち $I = 2$ の場合 Similarity Join Query Search と呼ばれる[11]．2 次元 ℓ_1 球は菱形であるので図 3 に示すような中心 $(0,0),(2W,0),\ldots$ で半径 W の ℓ_1 球で時刻空間 $\boldsymbol{T}$ を被覆できることを利用する．ここで菱形の個数は $(T_1T_2)/(2W^2)$ で与えられ式 (17) で述べた内接する正方形を用いる場合よりも少なくなり効率が良い．したがって式 (23) の距離計算回数比は式 (25) で与えられる．

$$\rho(W) = \frac{1}{a(W)}(2W^2). \tag{25}$$

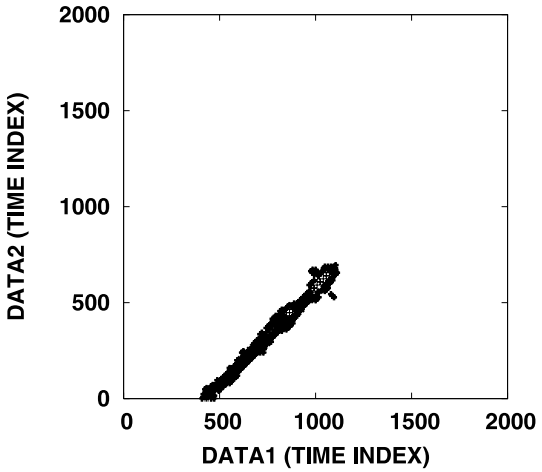
再分割は中心を上下左右に $W/2$ ずらし半径を $W/2$ とした 4 つの球であり，新たな球の中心に対して距離を求め式 (16) の判定を行う．W の初期値を式 (14) を満たす $W = 256$ とする．

図 5 に 2 つの時系列に対する RDDS 法の適用結果を示す．ここで $\theta = 0.2$，$W_{\min} = 8$ とした．図で x 軸が第 1 回目，y 軸が第 2 回目のデータであり，各軸の数字は時刻 t である．図 5 (a) から 2 つのデータ中で 4 カ所の類似セグメントが検出されている．図 5 (b) および (c) は検出した 4 カ所の中の第 1 番目のセグメントを拡大して表示している．図 5 (b) の時刻の組 $(408,0)$ 付近から延びる領域は検出された時刻で第 1 回目データの開始 6.5 秒目（408×16 ms），第 2 回目データの開始 0 秒から約 11 秒間に類似の系列を含んでいることを示している[⋆1]．図 5 (c) の領域は枝刈りの様子を示している．十字印の位置で距離計算の時刻を，十字の大きさで半径 W の大きさ（見やすさのため $W/2$）を表示している．検出/枝刈りの境界では判定が難しいので再帰が深く呼び出され W が小さくなっている．十字印が大きい領域は 1 回の距離計算で

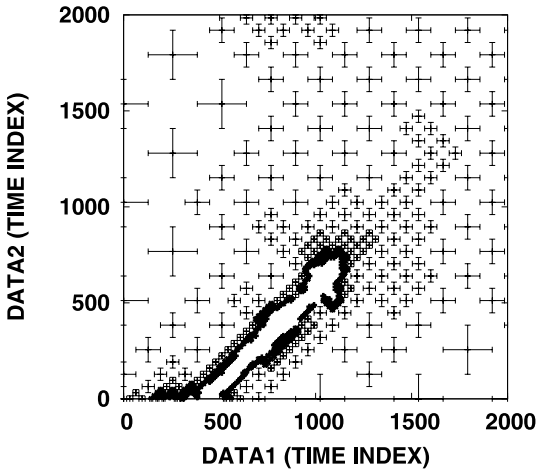
⋆1 このデータの発話内容は「周波数 76.2 MHz，出力 10 W で会津若松市中町からお送りします」である．



(a) 検出個所（4 カ所）



(b) 検出領域（第 1 番目のセグメント）



(c) 枝刈り領域（第 1 番目のセグメント）

図 5 RDDS 法で得られる検出および枝刈り領域（$\theta = 0.2$，$W = 256$，$W_{\min} = 8$）
Fig. 5 Detection and pruning regions in RDDS algorithm.

大きな菱形領域が枝刈り判定されている．θ をより小さく，もしくは $W_{\min}$ をより大きくすれば探索は高速になるが検出個所は減少する可能性がある．これらの最適値の自動決定は今後の検討課題である．

約 1 時間の 2 つの時系列に対して処理時間は約 1.432 秒である．VQ に 0.641 秒，出現確率計算に 0.284 秒，探索処理に 0.453 秒，データ読み込みなどに 0.054 秒である[⋆1]．AS 反復法での探索時間は約 74.080 秒であるので RDDS 法は約 163 倍（= 74.080/0.453）の高速動作が得られた．RIFAS 法は AS 反復法に比べて 20–40 倍程度の高速化が得られたと報告されているので RDDS 法は RIFAS 法の 4–8 倍の高速化が得られたことになる．1 章で指摘した菱形の上半面しか利用できていない RIFAS 法の欠点を改善できたことになる．$W = 256$ のとき，$a(W) = 1.882$ であるので $\rho(W) = 6.965 \times 10^4$ であり，式 (25) の ρ の上限値は $2L^2 = 7.813 \times 10^5$ となる．一方 AS 反復法の全数探索法に対する距離計算回数比は 3.100×10^2 である．この値は式 (3) のスキップ幅の平均値に対応している．$\rho(W)$ との比 $2.246 \times 10^2 = (6.965 \times 10^4)/(3.100 \times 10^2)$ が探索処理時間の改善比 163 倍に対応している．中心間の距離行列を利用することで探索時間の効率化を図ることが可能である[12),13)]．また 2 個の時系列を同時に保持するのではなく時系列 1 に対して時系列 2 を W もしくは $2W$ だけ読み込み処理をすれば少ない記憶容量で動作可能である．

4.3 3 つの時系列中の類似セグメント探索

3 つの時系列中の類似セグメント探索問題における RDDS 法の有効性を全数探索法を基準として距離計算回数および探索処理時間において評価する．AS 反復法を適用することも可能であるが，4.2 節で述べたとおり，AS 反復法は全数探索法と比較して 3.100×10^2 倍程度の改善が得られるだけである．3 次元空間の ℓ_1 球は正八面体であり，それを用いて 3 次元空間で式 (15) を満たす効率良い被覆方法は知られていないので式 (17) で述べたとおり**図 6** に示す 3 次元空間の ℓ_∞ 球である立方体を利用して 3 次元空間を被覆する．正八面体の半径を W とすると，式 (17) より，ℓ_∞ 球の半径は $W/3$ であり，立方体 1 辺の長さは $2W/3$ となる．式 (14) から初期半径 W は $(IL)/2 = (3 \times 625)/2 = 937$ 以下となる．初期半径 W をこれ以上大きくすると式 (13) が必ず成立しないので再帰分割が発生し，距離計算回数が増加することになる．ℓ_∞ 球の半径が奇数のとき，再帰分割において不足領域を生ずる[14)]ので，$W \le (IL)/2$ を満たし 3×2^k の整数，すなわち $W = 768 (= 3 \times 2^8)$ と設定する．

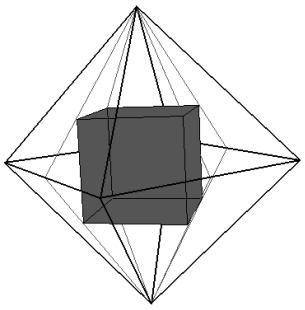


図 6 正八面体に内接する立方体
Fig. 6 Cube inscribed in an octahedron.

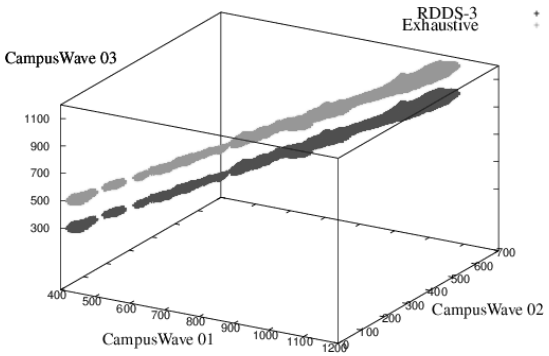


図 7 RDDS 法と全数探索法との探索結果の比較
Fig. 7 Comparison of search results with RDDS and full search algorithms.

4.3.1 60 秒データを用いた探索実験

1 時間の音声で全数探索を行うと膨大な時間を要するので，CampusWave データベースの最初の 60 秒を用いて評価する．**図 7** に RDDS 法と全数探索法の探索結果を示す．図で上が全数探索法，下が RDDS 法である．両方の探索結果は完全に一致しているが，2 つの探索結果を比較するため全数探索法の探索結果を上方にずらして表示している．この図から RDDS 法が全数探索法と同一の探索結果を与えることが分かる．**表 2** に全数探索法の探索処理時間と距離計算回数を示す[⋆2]．**表 3** に RDDS 法の探索処理時間と距離計算回数および全数探索法を基準とした改善比 $\rho(W)$ を示す．改善比は値が大きいほど，RDDS 法が高速探索処理であることを表す．

表 3 の全数探索法の点の組合せ数は，60 秒（$T =$

⋆1 計測には Pentium 4（2.66 GHz），メモリ 512 MB で OS Vine Linux 4.1 を載せたコンピュータを用い，10 回動作させた平均値である．

⋆2 計測には Pentium D 940（3.20 GHz），メモリ 2.0 GB で OS Vine Linux 3.2 を載せたコンピュータを用いた．

表 2 全数探索法の探索処理時間，距離計算回数
Table 2 Search processing time and number of distance calculations with full search algorithm.

探索処理時間 (sec)	距離計算回数
6,786	3.05×10^{10}

表 3 探索処理時間，距離計算回数における全数探索法からの RDDS 法の改善比
Table 3 Improvement ratio of RDDS from full search in search processing time and number of distance calculations.

	探索処理時間 (sec)		距離計算回数	
$W_{\min}$		改善比		改善比
0	1.37	4,953	3,126,082	9,771
1	0.11	61,690	305,284	100,060
2	0.03	226,200	96,292	317,231
4	0.01	678,600	41,732	731,977
8	0.01	678,600	25,536	1,196,228

3,126）であるので $T^3 = 3{,}126^3 = 3.055 \times 10^{10}$ となる．一方，正八面体の半径は $W = 768$ であるので立方体の 1 辺の大きさ $2W/3$ を用いてその個数は $(T/(2W/3))^3 \sim 6.105^3 = 227$ となり，再帰計算で得られる距離計算回数は 3,125,855 (= 3,126,082 − 227) となる．したがって再帰分割での距離計算回数の割合は 99.99% と非常に大きい．またこのときの平均距離計算回数は $a(W) = 3{,}126{,}082/227 = 1.377 \times 10^4$ となる．

4.3.2 30 分データを用いた探索実験

CampusWave データベースの第 1–3 回の各々の最初の 30 分を用いて探索実験を行い評価する．30 分のデータに対する全数探索法の探索処理時間は膨大であるのでここでは RDDS 法のみの評価を行う．全数探索法の距離計算回数，探索処理時間は時系列の長さ T の 3 乗に比例するので，30 分の音声長に対する距離計算回数および探索処理時間は次のように推定される．

(1) 距離計算回数：$(3.055 \times 10^{10}) \times 30^3 = 8.248 \times 10^{14}$

(2) 探索処理時間：6,786 (sec) $\times 30^3 = 1.832 \times 10^8$ (sec) $\sim$ 2,120 (days) = 5.808 (years)

図 8 に RDDS 法の探索結果を示す．各軸は 3 つの時系列の 30 分の時間長に対応している．一番左の点 $(0, 0, 0)$ 付近の探索結果は，図 7 の探索結果に対応している．**表 4** に RDDS 法の探索性能を示す．$W_{\min} = 0$ のときの探索処理時間および距離計算回数の改善比は 3.428×10^6，4.910×10^6 であり，4.3.1 項の実験における改善比 4,953 および 9,771 に比べて大きく上昇した．これは時系列における類似部分の占める割合が大きく違うことが理由である．$\theta = 0.3$，$W_{\min} = 0$ での

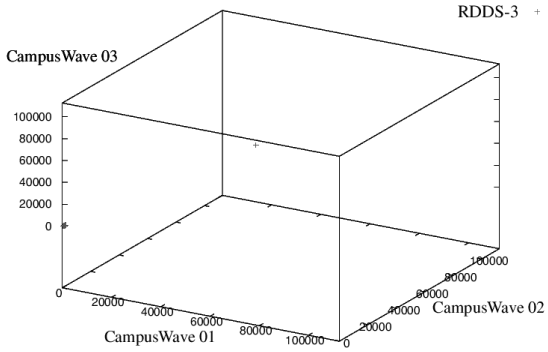


図 8 RDDS 法の探索結果（30 分データ）
Fig. 8 Search results with RDDS algorithm (30min data).

表 4 探索処理時間，距離計算回数における全数探索法からの RDDS 法の改善比
Table 4 Improvement ratio of RDDS from full search in search processing time and number of distance calculations.

閾値	最小半径	探索処理時間		距離計算回数	
θ	$W_{\min}$	(sec)	改善比 $\times 10^6$	$\times 10^8$	改善比 $\times 10^6$
0.3	0	53.45	3.428	1.679	4.910
	1	52.24	3.507	1.651	4.994
	2	52.05	3.520	1.649	5.000
0.6	0	141.09	1.299	3.705	2.226
	1	80.11	2.287	2.562	3.220
	2	76.65	2.390	2.478	3.329
0.9	0	257.33	0.712	67.11	1.230
	1	127.11	1.441	4.118	2.003
	2	119.36	1.535	3.923	2.102

全組合せにおける類似部分の割合は，4.3.1 項の実験では 2.84%であるのに対してこの実験では 0.01% である．RDDS 法は式 (13) でスキップ半径を導くが，類似部分付近でスキップ半径が減少して探索処理時間や距離計算回数を増加させるのに対して，それ以外の場所ではスキップ半径が大きくなる．一方，表 3 に示したように $W_{\min}$ の値を変化させることによって探索処理速度や距離計算回数を改善することができると予想したが，類似部分が少ない場合は改善効果は大きくないことが分かった．探索閾値 θ を 0.6，0.9 と大きくすると $\theta = 0.3$ の場合に比べて該当する類似部分が増加するため，探索処理時間および距離計算回数が増加する．この場合には $W_{\min}$ の値を 1，2 と増加させることによって改善することができることが分かる．30 分 ($T = 111{,}876 = 30\,\mathrm{min} \times 60 \times 10^3\,\mathrm{ms}/16 - L + 1$) に対する全数探索法の点の組合せの数は $T^3 = 1.400 \times 10^{15}$ である．3.4 節で述べたように $T/(2 * W/3) = 111{,}876/(2 \times 768/3) = 218.507$ であるので初期立方体の個数は $218.507^3 \sim 10{,}432{,}800 = 1.043 \times 10^7$

となる．$\theta = 0.3$, $W_{\min} = 0$ の表 4 の距離計算回数における再帰による距離計算回数は 15.747×10^7 $(= 1.679 \times 10^8 - 1.043 \times 10^7)$ となる．4.3.1 項の実験結果と異なり，再帰計算回数の割合は 93.788% と，減少していることが分かる．これは類似部分が少ないため立方体の再帰分割なしで探索判定が完了するものが増加していることを示している．平均距離計算回数は $a(W) = 16.098 = 1.679 \times 10^8 / 1.043 \times 10^7$ となり，4.3.1 項における値よりも減少することが分かる．

5. むすび

任意の複数時系列中の類似セグメント探索問題を定式化しスキップ半径の評価式を導出した．類似セグメント探索問題の解法として RDDS 法を提案し，2 つおよび 3 つのの時系列中の類似セグメント探索問題に適用しその有効性を示した．本論文では割愛したが時系列中のクエリ探索における有効性もすでに報告されている[13]．今後は計算量の理論的実験的な解析，探索閾値 θ，球の初期半径の最適値 $W_{\rm opt}$ および最小値 $W_{\min}$ の最適値の自動決定法，任意数の時系列中の類似セグメント探索の実装と評価，画像クエリ探索やクエリ集合探索問題の実装と評価，インタラクティブ探索表示システムの構築，より堅牢な特徴量や探索を効率化する VQ 符号帳の設計法，2 つの時系列中の類似セグメント探索法に帰着させる方法について検討する．

謝辞 ご討論いただき有益な助言をいただいた西村拓一博士（産総研）に感謝します．また 4.3 節の実験を行った酒井章裕君，日頃討論してくれたヒューマンインタフェース学講座の渡辺善之君，論文原稿を推敲し適切な助言をくれた渡邉括行君をはじめとする諸氏に感謝します．

参考文献

1) 柏野邦夫ほか：ヒストグラム特徴系列に基づく長時間音響信号の高速探索，音学講論，2-9-24, pp.561–562 (1998-09).
2) 杉山雅英：Active 探索におけるノルムと類似度との関係，音声研資，SP2004-173, pp.53–58 (2005-03).
3) 杉山雅英：セグメントの高速探索法，音声研資，SP98-141, pp.39–45 (1999-02).
4) 西村拓一ほか：時系列パターンの検索手法，信学技報，PRMU99-125, pp.173–180 (1999-11).
5) 西村拓一ほか：アクティブ探索法による時系列データ中の一致区間検出 —参照区間自由時系列アクティブ探索法，信学論 D-II，Vol.J84-D-II, No.8, pp.1826–1837 (2001).
6) 柏野邦夫ほか：二つの音響信号に共通に現れる部分信号区間の高速自動抽出，音学講論，1-Q-1, pp.133–134 (2000-3).
7) 杉山雅英：複数時系列中の類似セグメント探索法の提案，音学講論，1-1-9, pp.17–18 (2006-03).
8) 杉山雅英：セグメント探索のためのノルム及び集合の諸性質，情報処理学会東北支部第 1 回研究会，15 (2005-12).
9) 杉山雅英：距離に基づく Active 探索法の計算量について，音声研資，SP2006-9, pp.19–24 (2006-06).
10) 内田貴文，杉山雅英：CampusWave 音声データベースの作成，電気関係学会東北支部連合大会，2A-6 (2000-08).
11) Dohnal, V.: Indexing Structures for Searching in Metric Spaces, Ph.D. Thesis, Masaryk University (Feb. 2004).
12) 杉山雅英，岡本知子：距離空間と出現確率時系列の幾何学的性質に基づくセグメント高速探索法，情報処理論文誌，Vol.47, No.6, pp.1675–1686 (2006).
13) 渡辺善之ほか：類似セグメント探索 RDDS 法の評価，音声研資，SP2006-99, pp.83–88 (2006-12).
14) 杉山雅英：類似セグメント高速探索法における球被覆の検討，音学講論，1-P-22 (2007-03).

付　録

d を ℓ_p 距離とし，ρ（≥ 1）に一般化した距離を以下のように定義する．

$$d_\rho(\boldsymbol{t}) = \begin{cases} \left(\sum_{i<j} d^\rho(t_i, t_j) \right)^{\frac{1}{\rho}}, & (\rho \geq 1) \\ \max_{1 \leq i < j \leq I} d(t_i, t_j), & (\rho = \infty). \end{cases}$$

性質 1 は $\rho = 1$ の場合に該当する．

A.1 性質 1 の証明

性質 3 (1) 単調減少性

$1 \leq \rho < \sigma \Rightarrow d_\sigma(\boldsymbol{t}) \leq d_\rho(\boldsymbol{t})$.

(2) $\boldsymbol{n} = (n_i)$ に対して $\hat{\boldsymbol{n}} = \left(\frac{|n_i|}{L_i} \right)$ とするとき，

$$|d_\rho(\boldsymbol{t} + \boldsymbol{n}) - d_\rho(\boldsymbol{t})| \leq 2^{\frac{1}{\rho}} c_{I,\rho} \, ||\hat{\boldsymbol{n}}||_\rho. \qquad (26)$$

ここで $c_{I,\rho} = 2 \left(\frac{I-1}{2} \right)^{\frac{1}{\rho}}$ $(c_{I,1} = I - 1)$ である．

証明：(i, j) を添字とするベクトルを $\boldsymbol{a} = (d(t_i, t_j))$ とすると $d_\rho(\boldsymbol{t}) = ||\boldsymbol{a}||_\rho$ となる．ノルムの単調減少性から (1) は明らか．$\boldsymbol{b} = (d(\boldsymbol{t}+\boldsymbol{n}))$ とすると $d_\rho(\boldsymbol{t}+\boldsymbol{n}) = ||\boldsymbol{b}||_\rho$ となるので $|d_\rho(\boldsymbol{t}+\boldsymbol{n}) - d_\rho(\boldsymbol{t})| = |\, ||\boldsymbol{b}||_\rho - ||\boldsymbol{a}||_\rho \,|$ となる．ノルムの三角不等式から $|\, ||\boldsymbol{b}||_\rho - ||\boldsymbol{a}||_\rho \,| \leq ||\boldsymbol{b} - \boldsymbol{a}||_\rho$ を得る．$\boldsymbol{b} - \boldsymbol{a} = (d(t_i + n_i, t_j + n_j) - d(t_i, t_j))$ であるので

$$||\boldsymbol{b}-\boldsymbol{a}||_\rho = \left(\sum_{i<j}|d(t_i+n_i,t_j+n_j)-d(t_i,t_j)|^\rho\right)^{\frac{1}{\rho}}.$$

補題 1 から絶対値部が評価されるので以下を得る．

$$||\boldsymbol{b}-\boldsymbol{a}||_\rho \le 2^{\frac{1}{\rho}}\left(\sum_{i<j}\left(\frac{|n_i|}{L_i}+\frac{|n_j|}{L_j}\right)^\rho\right)^{\frac{1}{\rho}}.$$

$\hat{\boldsymbol{n}} = \left(\frac{|n_i|}{L_i}\right)$ に補題 2 を適用して $\le 2^{\frac{1}{\rho}}c_{I,\rho}||\hat{\boldsymbol{n}}||_\rho$ を得る．

補題 1 以下の不等式が成り立つ．

$$|d(t_i+n_i,t_j+n_j)-d(t_i,t_j)| \le 2^{\frac{1}{\rho}}\left(\frac{|n_i|}{L_i}+\frac{|n_j|}{L_j}\right).$$

証明：$|d(t_i+n_i,t_j+n_j)-d(t_i,t_j)|$
$\le |d(t_i+n_i,t_j+n_j)-d(t_i,t_j+n_j)|$
$+|d(t_i,t_j+n_j)-d(t_i,t_j)|$.
ここで第 1 項および第 2 項の時刻 t_j+n_j および時刻 t_i は一定であるので式 (1) から以下を得る．

$$|d(t_i+n_i,t_j+n_j)-d(t_i,t_j)| \le \frac{2^{\frac{1}{\rho}}}{L_i}|n_i|+\frac{2^{\frac{1}{\rho}}}{L_j}|n_j|.$$

補題 2 $\rho \ge 1$ と n 個の実数 $a_i \ge 0$ に対し以下が成り立つ．

$$\left(\sum_{i<j}(a_i+a_j)^\rho\right)^{\frac{1}{\rho}} \le c_{n,\rho}\left(\sum_{i=1}^n a_i^\rho\right)^{\frac{1}{\rho}}.$$

ここで $c_{n,\rho} = 2\left(\frac{n-1}{2}\right)^{\frac{1}{\rho}}$.
証明：右辺の $(\sum_{i=1}^n a_i^\rho)^{\frac{1}{\rho}} = ||\boldsymbol{a}||_\rho = 0$ の場合は $a_i = 0$ $(\forall i)$ であるので左辺も 0 となり，不等式が成り立つのは自明．したがって右辺 $||\boldsymbol{a}||_\rho \ne 0$ として，両辺を $||\boldsymbol{a}||_\rho$ で割り $\hat{a}_i = a_i/||\boldsymbol{a}||_\rho$ をおくと $(\sum_{i<j}(\hat{a}_i+\hat{a}_j)^\rho)^{\frac{1}{\rho}}$ となる．$||\hat{\boldsymbol{a}}||_\rho = 1$ であり有界閉集合上の連続関数は最大最小値を持つので左辺は最大値 $c_{n,\rho}$ (>0) を持つ．次に $c_{n,\rho}$ を定める．${}_n\mathrm{C}_2$ 個の $(a_i+a_j)^\rho$ の相加相乗平均の関係より $(\sum_{i<j}(a_i+a_j)^\rho)^{\frac{1}{\rho}} \le ({}_n\mathrm{C}_2(\prod_{i<j}(a_i+a_j)^\rho)^{\frac{1}{{}_n\mathrm{C}_2}})^{\frac{1}{\rho}} = {}_n\mathrm{C}_2^{\frac{1}{\rho}}\prod_{i<j}(a_i+a_j)^{\frac{1}{{}_n\mathrm{C}_2}}$．$a_i+a_j \le 2(a_ia_j)^{\frac{1}{2}}$ より $\le ({}_n\mathrm{C}_2)^{\frac{1}{\rho}}\prod_{i<j}2^{\frac{1}{{}_n\mathrm{C}_2}}(a_ia_j)^{\frac{1}{2}\frac{1}{{}_n\mathrm{C}_2}} = 2({}_n\mathrm{C}_2)^{\frac{1}{\rho}}(\prod_{i=1}^n a_i)^{\frac{1}{n}}$．ここで Lagrange の未定乗数法を用いて正の数列の和が一定値のとき，数列の積は数列が等しい，すなわち，$a_i = n^{-\frac{1}{\rho}}$ のとき，最大となるので $\le 2\left(\frac{n-1}{2}\right)^{\frac{1}{\rho}}$．一方，$a_i = n^{-\frac{1}{\rho}}$ のとき，$(\sum_{i<j}(a_i+a_j)^\rho)^{\frac{1}{\rho}} = 2\left(\frac{n-1}{2}\right)^{\frac{1}{\rho}}$ であるので証明完了．

A.2 性質 2 の証明

平均化窓長 L_i が等しいとは限らない場合の式 (12), (13) を一般化して不等式で証明する．時刻 $\boldsymbol{t}$，$\boldsymbol{t}+\boldsymbol{n}$ $(\boldsymbol{t}=(t_i), \boldsymbol{n}=(n_i))$ における距離 $d_\rho(\boldsymbol{t})$ の式 (26) の右辺を平均化窓長 $\boldsymbol{L} = (L_i)$ の重み付けノルム $||\boldsymbol{n}||_{\rho,\boldsymbol{L}} = \left(\sum_i\left(\frac{|n_i|}{L_i}\right)^\rho\right)^{\frac{1}{\rho}}$ で表わす．

$$|d_\rho(\boldsymbol{t}+\boldsymbol{n})-d_\rho(\boldsymbol{t})| \le 2^{\frac{1}{\rho}}c_{I,\rho}||\boldsymbol{n}||_{\rho,\boldsymbol{L}}$$
$$-2^{\frac{1}{\rho}}c_{I,\rho}||\boldsymbol{n}||_{\rho,\boldsymbol{L}} \le d_\rho(\boldsymbol{t}+\boldsymbol{n})-d_\rho(\boldsymbol{t})$$
$$\le 2^{\frac{1}{\rho}}c_{I,\rho}||\boldsymbol{n}||_{\rho,\boldsymbol{L}}$$

であるので $||\boldsymbol{n}||_{\rho,\boldsymbol{L}} < |\theta-d_\rho(\boldsymbol{t})|/(2^{\frac{1}{\rho}}c_{I,\rho})$ が成り立つとき，$d_\rho(\boldsymbol{t}) > \theta$ であれば $d_\rho(\boldsymbol{t}+\boldsymbol{n}) > \theta$，$d_\rho(\boldsymbol{t}) < \theta$ であれば $d_\rho(\boldsymbol{t}+\boldsymbol{n}) < \theta$ が成り立つ．証明完了．

(平成 19 年 4 月 3 日受付)
(平成 19 年 10 月 2 日採録)

杉山 雅英（正会員）

1954 年生．1977 年東北大学理学部数学科卒業．1979 年同大学院理学研究科数学専攻修士課程修了．同年日本電信電話公社武蔵野電気通信研究所（現 NTT 武蔵野研究センター）入所．1985 年東北大学より工学博士号を取得．1986 年から米国 AT&T Bell 研究所滞在研究員，1987 年から NTT 基礎研究所主任研究員，1990 年から ATR 自動翻訳電話研究所主幹研究員の後，1993 年から会津大学コンピューター理工学部ヒューマンインタフェース学講座教授．現在まで音声特徴キーによる音声検索等の音声認識処理の研究に従事．日本音響学会，電子情報通信学会，IEEE 各会員．